

# COST, CONTEXT, AND DECISIONS IN HEALTH ECONOMICS AND HEALTH TECHNOLOGY ASSESSMENT

Anthony J. Culyer

University of York - Economics & Related Studies [tony.culyer@york.ac.uk](mailto:tony.culyer@york.ac.uk)

**Objectives:** This study is an attempt to demystify and clarify the idea of cost in health economics and health technology assessment (HTA).

**Methods:** Its method draws on standard concepts in economics. Cost is a more elusive concept than is commonly thought and can be particularly elusive in multidisciplinary territory like HTA.

**Results:** The article explains that cost is more completely defined as opportunity cost, why cost is necessarily associated with a decision, and that it will always vary according to the context of that decision: whether choice is about inputs or outputs, what the alternatives are, the timing of the consequences of the decision, the nature of the commitment to which a decision maker is committed, who the decision maker is, and the constraints and discretion limiting or liberating the decision maker. Distinctions between short and long runs and between fixed and variable inputs are matters of choice, not technology, and are similarly context-dependent. Harms or negative consequences are, in general, not costs. Whether so-called “clinically unrelated” future costs and benefits should be counted in current decisions again depends on context.

**Conclusions:** The costs of entire health programs are context-dependent, relating to planned rates of activity, volumes, and timings. The implications for the methods of HTA are different in the contexts of low- and middle-income countries compared with high-income countries, and further differ contextually according to the budget constraints (fixed or variable) facing decision makers.

**Keywords:** Opportunity cost, Context, Short/long runs, Fixed/variable costs, Unrelated costs, LMICs

“Obscurities, ambiguities, and errors exist in cost and supply analysis despite, or because of, the immense literature on the subject” (1, p 273). What Alchian could write in 1959, when his article was first published, is apt in today’s health technology assessment (HTA) (and health economics in general). The root idea of “cost” is “opportunity cost” but this idea is ill-understood (or, better, differently understood) not only by people who are not economists, but even by economists including, we may presume, health economists who work in HTA.

This study is an attempt to demystify and clarify the idea of cost, not only for health economists but also for HTA analysts from other disciplines and for members of multidisciplinary teams working on economic evaluations in the field of health, health care, and social care. This basic idea, that costs (and, for that matter, benefits) cannot be defined without specifying the context of the decision, turns out to be a recurring theme. I shall assume, to avoid residual ambiguity, that the costs with which we are concerned in decisions about healthcare investments (using HTA) are the costs of resources and that the opportunities forgone are those arising from the varied alternative uses to which those resources might be put.

I am grateful for the comments of three anonymous reviewers and for discussions with and comments from Kalipso Chalkidou, Karl Claxton, Richard Cookson, Mike Drummond, Ali McGuire, Alec Morton, Mark Pauly, Adrian Towse, and Platonas Yfantopoulos.

## Confusion, Even Among Economists

In 2005, Ferraro and Taylor (2) put a question based on one by Frank and Bernanke (3) to 199 economists. I have adapted it by giving it a health context<sup>1</sup>:

*You have a health condition that could be treated without charge by your public national health insurance programme. You could also “go private” by registering with a local surgical clinic offering medical care of an equivalent quality for which you would pay out of pocket, which is your next-best alternative choice. The price of private treatment is \$400 and the maximum you would be willing to pay for private care is \$500. Assume there are no other costs of either option. Based on this information, what is the opportunity cost of choosing public health care:*

- A. \$0
- B. \$100
- C. \$400
- D. \$500?

<sup>1</sup> The test question in the original was: Please Circle the Best Answer to the Following Question:

You won a free ticket to see an Eric Clapton concert (which has no resale value). Bob Dylan is performing on the same night and is your next-best alternative activity. Tickets to see Dylan cost \$40. On any given day, you would be willing to pay up to \$50 to see Dylan. Assume there are no other costs of seeing either performer. Based on this information, what is the opportunity cost of seeing Eric Clapton?

- A. \$0
- B. \$10
- C. \$40
- D. \$50

The most popular answer was \$500, with 27.6 percent of respondents choosing this answer. The second most popular answer was \$400, with 25.6 percent of respondents choosing it. The third was \$0, with 25.1 percent of respondents choosing it. The answer \$100 was the least popular, with only 21.6 percent of respondents choosing it. The answers from these well-trained economists seemed to be randomly distributed across the four possible answers. What is your answer? What is the correct answer?

According to Ferraro and Taylor, \$100 was the right answer: the value to you of the forgone care (\$500) less the \$400 out-of-pocket price you also forgo. In this case, it is essentially what economists call the “consumer’s surplus” forgone. But you could have taken a different view: you could have valued the forgone opportunity at its market price (\$400), which one may suppose represents the minimum sum the providers must receive for their services. Or you might have taken the view that the opportunity cost is best considered descriptively: what was forgone was private care with a bundle of associated characteristics which you say you value at \$500. Or you might say the opportunity cost was \$0, because whatever you decide, the service will be provided to someone with no additional resource costs than if it were provided to you.

This kind of confusion is plainly a serious matter, because opportunity cost is one of the half dozen key economic concepts whose understanding defines a professional health economist. One reason for confusion in this case may have been the unfamiliar context: the experiment was not framed in terms of the opportunity cost of resources used in the production of a good or service but rather in the context of choice between a final good or service (see Parkin for other explanations) (4). The \$100 opportunity cost is subjective and is revealed neither by the prices of the options nor in any system of accounting. The other two interpretations (C and D) follow straightforwardly: the market price of the private care (\$400) on the one hand and the general characteristics of the forgone option (choice of physician, early admission, etc.) valued at \$500 on the other.

From a public decision-maker’s point of view the most useful approach may be any one of these: the first to understand patient behavior and the (dis)incentives required to change it, the second to appreciate what must be forgone in the health care system for an expenditure of \$400 per patient on a public service contracted to a private service supplier, the third to discover whether a publicly provided service package might be provided at a lower cost than a contracted-out service having somewhat different characteristics and the fourth simply to dismiss the whole issue as a question of transfers rather than costs.

All four answers thus may have something to be said for them as a reasonable interpretation according to context. What is reasonable does seem to depend on context: what is the question, who is the decision maker, and what is the

decision maker seeking to achieve? It should not be surprising that opportunity cost will vary according to identity (e.g., position in a hierarchy of decision makers) and the objectives of the decision maker (utility maximizing, profit maximizing, health maximizing ...).

#### Opportunity Cost and the Production of Services

Cost is rooted in the apparently simple idea that resources are best allocated in ways that ensure that they wind up in their most valued uses: uses that are more valued than alternatives that might have been chosen instead. Early emphasis on the importance of opportunity cost in economics is particularly associated with the Austrian School of economics in the late nineteenth century:

*To say that any kind of production involves cost simply implies that the economic means of production, which could doubtless have been usefully employed in other directions, are either used up in it, or are suspended in it. (5, p 175)*

This is what Robbins (6, p 2) described as Wieser’s law. “Cost” in economics is a term applied to the value of resources, most commonly “factors of production,” that have alternative uses. The term “opportunity cost” may refer to the value of the most preferred alternative use to the current or a planned use; or it may refer to the quantities of a good or service that is forgone, like an alternative drug, or an alternative nurse; or it may refer to an alternative nonmonetary measure of value, most notably for health economists, health itself, so that the opportunity cost of a decision becomes the health that would have been generated by the resources in question had the decision gone in favor of the preferred alternative. Whichever of these it may be, cost is always the value of the best alternative necessarily forsaken when resources are committed, resources that could have been put to many other uses inside or outside the healthcare sector.

In a well-informed market with low transaction costs, the value of the forgone opportunity (say, use of a resource) may be objectively revealed as a dollar amount, representing the highest value someone else places upon it as revealed in offers they make, and where they will need to be outbid if several are bidding, or compensated if they already own the resource and are to make it available to a bidder in a notional auction. Evidently, markets rarely operate in such a perfect manner and so prices may depart from the “true” opportunity cost, how much will depend on a judgment by experts familiar with that market.

Within an organization like a hospital or an ambulance service, resources also have alternative uses, some of which are within the organization (e.g., rooms can be reallocated to other uses) and some of which lie outside (e.g., ambulances can be sold to other users). The external uses may reveal opportunity cost as a monetary price. The internal opportunity cost is not, however, revealed in a market price but is subjective,

existing only in the mind of the decision maker (7;8). This is not to say that the subjective judgment that a particular resource is best used in a particular way is not a communicable judgment. Plainly that is not the case, as managers, who often have to act upon their subjective valuations of opportunity cost, can be held to account for their decisions through audit trails, direct enquiries, and so on, and can give their reasons for making the decisions they made when called to account.

Cost is always the consequence of a decision and is in turn a consequence of the resource scarcity that standard texts tell us (at least since Robbins) (9) lies at the heart of economics as an analytical discipline. All economists seem to agree that the forgoing of a want or need that could be satisfied through the use of resources is fundamental and not itself contextual. It is what *defines* cost. All else, however, is contextual. How best the forgone option is measured or valued will, therefore, depend on who the decision maker is and the question the decision maker wants answered.

#### Costs versus Undesired Attributes

A significant implication of the idea of opportunity cost is that negative consequences are not the same as costs. There is a tendency by the theoretically naive, when applying techniques like multi-criteria decision analysis, to create categories (often overlapping) of benefits and costs treated broadly as “pros” and “cons” (10). Loose usage easily transforms “harm” or “negative consequence” into “cost”. Consider the homely example given by Alchian, whose extensive reworking of the cost theories of the Austrian School (notably von Wieser and the early English marginalists) (5;11–13) is drawn on extensively here.

Suppose you want to build a pool in your garden. “The toil and trouble of digging it and the nuisance of noisy disobedient neighborhood children and uninvited guests who use it are undesirable attributes of the pool. They are not the costs of creating and having the pool. This distinction between (a) undesirable attributes inherent in some event and (b) the highest-value forsaken option necessary to realize the event is fundamental, for only the latter is cost as the term is used in economics” (1, p 301–302). Blood, sweat, and tears associated with building the pool are not costs. If they are your tears, they may be relevant to the decision in two ways: they reduce the value of creating and having the pool and they may be, indeed probably are, attributes of some of the other human inputs in the pool-building project which may need compensation if these resources are to be made available and thereby enter into costs. But they are not costs of building the pool because they are not resources having alternative uses. If, however, you are doing any of the digging, then your time is of course a resource with alternative uses and is accordingly a cost.

#### Cost and Decisions

Cost is the value of an opportunity for using a resource that is necessarily forsaken when a decision is made. The character of

the decision in question will normally affect its cost. This suggests (as we have already seen) that what is considered to be a cost is context-dependent.

As has again been seen, some costs are subjective, not readily revealed in systems of accounts, and only indirectly revealed by decisions. One aspect of context relates to the decision and the nature of the commitment being entered into. One may distinguish, for example, between a decision to acquire (acquisition cost) and a decision to operate a program of costs and outcomes over a period of time. Of course, one may combine them if the decision is both to acquire a resource and to use it in particular ways for a period. But one may also consider these as two (successive) decisions. Because the period over which the production program is to be run is also a matter of choice, there is an opportunity cost of each program according to its length. The cost of acquisition is a consequence of a decision to buy. Because it may not be a decision also to hold the acquired asset, the actual cost of buying is simply the immediate depreciation in value when the resource changes ownership: the difference between the purchase price and its immediate resale value.

The decision to operate a program also entails expenditures (usually called variable costs) that are incurred over time and whose present (discounted) value is the cost of running the program at the time of decision. The variable costs in health care, as elsewhere, are typically a function of the rate of activity, the planned volume of work, the date at which it is to begin, and the length of the period over which it is to run.

A typical healthcare program might embody both the acquisition and the continuing use of a resource, usually combined with others, over a planned period, with planned or anticipated rates of use by a target population. These are all matters for the decision maker to decide and different decisions (according to context) generate different costs and benefits. In HTA, it is usual to consider first the inherent potential of the resource viewed as an “intervention” or “technology.” One category of consideration is the expected direct health benefit per individual and any associated positive or negative health-related consequences as viewed by the decision maker. These would typically be discounted over the expected lifetime of the individual beneficiaries. The opportunity costs would be discounted over the period in which costs are incurred (the maximum being the expected remaining lifetimes of the beneficiaries).

A second category to consider is the intended population who will receive the intervention. The present value of the program costs of the decision to provide a technology of a specified type for a population of a specified type under circumstances of a specified type for a specified period then constitute the general opportunity cost of the decision to run the program, that is, the opportunity cost of the resources in their most preferred alternative uses, usually as revealed in a market, wherever those uses may be.

The present value of a cost or benefit accruing in the distant future is quite small. For example, at a 3 percent discount rate a dollar in 50 years' time has a present value of just above 2 cents and over 30 years of approximately 4 cents. At 8 percent, the present values are one cent in 30 years' time and one-fifth of a cent after 50 years. The importance of accuracy in distant costs and benefits is evidently less than that of closer costs and benefits.

#### Budget Constraints

Even a high-level healthcare decision maker is typically required, as a matter of either policy or managerial control (or both), to operate within a budget. This is as true of a private sector insurer or provider as it is of a public sector insurer or provider. There may be some tolerance around a specific annual planned budget, or some intertemporal financial substitution, but in general all decision makers need to optimize the use of limited resources. How they optimize their decisions is likely to depend on different criteria. A public sector payer might typically specify one criterion (among others) such as maximizing the impact of the available budget on population health. In such a case, the opportunity cost of investing in one program rather than another may be most usefully considered in terms of the forgone population health that the most preferred alternative use would have been expected to have generated. This is not subjective and can be empirically estimated (as, for example, in Claxton et al.) (14).

The determination of the budget is a matter for decision makers higher in the hierarchy of decision: for example, the senior management team and board of a commercial insurer, the senior management and trustees of a nonprofit payer, or a cabinet of ministers in a government. Below them, lower tier decision makers may also assign budgets to various departments, programs, and activities. In each case, the decision maker is appointed by and is accountable to a still higher authority, which is the ultimate source for the values of the organization as stated in, for example, its vision, aims, and objectives, and in any specific obligations, some of which may be legal requirements, placed upon the decision maker. The governance arrangements and lines of authority and accountability of a decision-making organization constrain decision makers similarly constrain the role of economic advisers, whose accountability is typically to the decision maker. HTA analysts are, of course, at liberty to state their own values, without any accountability other than to their own consciences, but such studies would be a species of advocacy unless there were a coincidental symmetry between these values and those of decision makers in the relevant sector.

#### Scope of Decisions

The program scope (whom covered, for what conditions, and over what period) is a matter of choice for the decision

maker. The scope of a decision also defines the perspective of an economic evaluation and, therefore, the alternative uses of resources being considered. The scope essentially specifies the matters of interest to the decision maker and generally includes such considerations as the technology or technologies to be evaluated and their comparators; the character of benefits to be assessed; the expenditures (budget impact) and costs to be considered; the types of individual whose interests as patients, family carers, clinical professionals, service providers, product manufacturers, and so on, are to be embraced; the measures of benefit; the periods over which costs and benefits are to be measured and any harms or nonhealth benefits that may accrue. These, by no means a complete list, all have the characteristic that they require social value judgements to be made about their inclusion or exclusion and the weight to be attached to each when they are combined. There are also technical specifications which do not embody social value judgements but which may be judgments of another kind, for example, judgments about what is good science or good evidence (15). Such judgments are sometimes embodied in statements of "best practice" like a reference case (16).

While defining scope and making social value judgments are the responsibility of decision makers, it does not follow that analysts have no role to play. On the contrary, they are by experience better at distinguishing between social value judgments, other kinds of value judgment (e.g., as to what is "good" science), and other kinds of judgment, both qualitative and quantitative. They can brief decision makers on the ethical implications of possible decisions, and they can seek logical consistency. Decision makers are apt to make some decisions for illegitimate reasons, or without weighing the alternatives. For example, in deciding the scope of a specific HTA, a decision maker might implicitly exclude some consequences through oversight, for political convenience, to avoid inconvenient conflicts, or through inability to imagine some of the collateral possibilities. The analyst should be able to alert decision makers to such possibilities so that the decision, when taken, can be seen to be a considered outcome of a process. This does not require the analyst to steal the decision maker's accountable role, but rather to support it in principled ways.

Scope is also contextual. Consider the perspective adopted in a low- or middle-income country (LMIC) by an internal decision maker in the Ministry of Health compared with that adopted by an external aid giver. At the aggregate level concerning the size, target population, and periodicity of the aid, the opportunity cost as seen by the aid giver might be seen exclusively in terms of alternative countries, populations, and dates while that seen by the internal decision maker will characteristically concern the consequences of substitution effects as domestic funding is replaced in part or whole by aid funding and domestic resources are released for other uses (including nonhealth uses such as better roads and more secondary education). Internally, the opportunity cost of declining

the aid includes the loss of a variety of health and nonhealth benefits of which only some may be of concern to the aid giver. The opportunity cost for each party is thus defined by its own objective function and budget constraint.

#### Related and Unrelated Costs

It is not customary in HTA to consider the costs of so-called “unrelated” costs and benefits. Morton et al. (17) briefly reviewed a long-standing controversy over the treatment of these effects, defined as consequences of a decision that prolongs life and, hence, may involve future costs (and benefits) that would not otherwise have been incurred. They make a strong argument for a balanced approach, that is, either include unrelated costs and benefits or exclude both. They argue, following several other contributions, that the correct approach is to include both (suitably adjusted of course for timing and uncertainty). Is this correct?

The first notable problem with the recommendation of Morton et al. is that it is vulnerable to an *ad nauseam* problem of infinite regress. It is one thing to consider the costs of future expected treatments that result from extensions to life but another to take account of the further life extensions that these future interventions may make possible, the new children who may be born who would not otherwise have been born, and who will doubtless have future healthcare needs. Indeed, if *all* unrelated costs are, as a matter of principle, to be considered, no matter how remote in time or geography, then there appears to be no limit to the costs to be considered other than those imposed by discounting and time preference on the one hand and the sheer cost of making the necessary estimates on the other. A solution might be pragmatic, with bounds on scope set explicitly but arbitrarily to avoid the infinite regress.

There is, however, a deeper problem with including unrelated costs (and benefits), understanding which enables more satisfactory guidance on the treatment of unrelated costs. The deeper problem arises from the insight by Morton et al. of official guidance on the conduct of economic appraisal in the public sector (18) as follows: “the test for including or excluding a cost or benefit depends on whether it flows from the commitment which we are being asked to make at the moment of decision.” That is indeed so. However, we need both to be satisfied that a consequence does actually flow from the decision and to enquire as to the nature of that “commitment.”

The idea of “commitment” suggests some form of contract, for example, one having the nature that an extended survival consequent to the current decision will entitle an individual to such further (clinically unrelated) interventions as may be deemed needed (or demanded). Such contracts cannot be taken for granted. In a healthcare system having a defined benefits package, future unrelated costs of interventions not in the package at the future date will not be incurred by the system’s

current decision maker and the present value of any expected expenditures made by others are, therefore, not costs of the current decision. By contrast, in an established system that is comprehensive in its coverage of people and treatments, the reasonable expectation is that there is a commitment to meet future consequential but unrelated costs, including those to be incurred by the unborn descendants of the present generation of beneficiaries.

Thus, it becomes clear that the treatment of unrelated costs is context-dependent. In LMICs only now venturing into a long-term strategy for universal health coverage, if the reasonable expectation is that future unrelated care needs will not be covered in the benefits package, then they should not be included as costs of current decisions, unlike the future costs of treating clinically related conditions (unless these are exempted by the decision maker as an explicit social value judgment). In countries where the expectation is that unrelated treatments will be delivered, or in any scheme in which future unrelated consequential care is provided for, then unrelated costs ought to be included in the costs of the current decision. In private health insurance schemes, whether or not unrelated future costs are properly included depends again on the contract, that is, what’s in and what’s out.

Benefits and costs should in general be treated symmetrically. If, for example, future unrelated costs are not covered by the contract, then estimates of future benefits that implicitly include the benefits of future unrelated interventions should be adjusted downward and, in the reverse case when unrelated treatments are included, estimates of the present value of subsequent benefits ought also to be included.

Some past advocates of inclusion (e.g., Morton et al.) (17) rightly take NICE to task when it prescribes in its Guide to the Methods of Technology Appraisal (19) “costs that are considered to be unrelated to the condition or technology of interest should be excluded.” On the contrary, the kind of implicit contract represented by the NHS describes precisely the circumstances under which such costs ought to be *included*. Use of life tables to estimate benefits from NICE-approved interventions almost certainly include the benefits of future unrelated interventions, which suggests further that NICE’s methodology is inconsistent by treating future unrelated costs and benefits asymmetrically.

#### Rates and Volumes

There is an ambiguity in the idea of an outcome or output in health care, as elsewhere, which can mislead. The ambiguity is this: are we speaking of rates (outputs per unit of time) or volume of output (cumulative outputs over time)? The relation between cost and rates or volumes in economics is in general qualitative, in the sense that the relationship can be signed as positive or negative. In formal modeling of cost functions, these characteristics are suggested by the structure and

parameters of equations. Beyond a general theoretical expectation that sooner or later marginal costs will rise with rising rates of output, *a priori* cost theory is silent on whether the effect size is positive or negative and substantial or not. It is entirely an empirical matter. At a pragmatic level, the following empirical generalizations may probably be deployed by HTA analysts with confidence and to good effect:

Implementation of a decision after a period for planning has a lower cost than immediate implementation. Reasons for this include improved information enhancing the effective use of medicines and other inputs and the possibility of lower future prices through negotiation. When a decision maker wishes to implement a decision very quickly (one might say “in the short run”), the present value of cost will be higher than otherwise. A classic example is the response to an epidemic. The case for advance stock-piling of medicines and fast-response teams of specialists is that they are relatively low-cost means of achieving the benefits of an emergency intervention. At the time of actually using these resources, however, the alternatives forgone at the time of their creation are sunk costs and quite irrelevant to the current choice situation, when the opportunity costs relate to the alternative uses of the stocks and teams rather than to their creation.

Notwithstanding the point about implementation, delays in implementation may be warranted when the estimated cost reductions or benefit enhancements exceed any losses of benefit attributable to the delay. Additional information is not worth collecting if the reverse is the case (20).

#### Sunk Costs, Fixed and Variable Costs

Sunk costs are costs that have been incurred by past decisions. They should generally be ignored in decision making, other than as learning experiences that may inform future decisions. The fact that resources were forgone in a decision that is already made is not a determinant of the current alternative use value of resources. Thus, the fact that large sums may have already been expended on a health project is not an argument for completing the project if the current assessment of the opportunity cost of completing it turns out to exceed the present value of the benefits of completion. The fact of heavy past expenditures may be deployed in a face-saving exercise but that represents an investment in something other than health or health care.

Distinctions between short and long runs are context-dependent and not for arbitrary measurement. A short run is one in which the decision maker *chooses* not to vary one or more inputs, that is, they are treated as fixed even though they could in principle be varied. The long run is when all, or at any rate a significant number of inputs, are chosen to be varied. Neither, despite their names, has anything to do with periods of time. Both are determined by an interaction between the desires of the decision maker and the costs of changing specific inputs, some of which may be changed

quickly with relatively low adjustment costs but others not. In other words, the decision maker chooses the “run,” it is not a technically fixed or exogenous variable.

Whether a cost is treated as fixed or variable is context-dependent. Thus, a contracted rental lease on a building or a contract of employment may be taken as given in many contexts involving the use of the resources under contract. Contracts can, however, always be renegotiated (at a cost) or breached (at a cost) and so become variable when decision makers judge the circumstances to warrant adjusting a fixity.

Resources that are apparently technically fixed are also variable in some contexts. Buildings and railway track are examples of resources that seem fixed for most purposes. However, both are subject to maintenance and minor alteration so that, over time, substantial increases or reductions in infrastructure are achieved. They can even be varied very quickly when it is so wished. When the English Great Western Railway’s old broad-gauge track was changed to the modern standard narrow gauge in 1892, the entire stretch of 213 miles from Exeter to Penzance was changed in one weekend (21). Thus, for some decisions (e.g., usual timetabling and ticketing) the track is assumed fixed and for others (e.g., system design) it is assumed variable. Moreover, 177 miles of the stretch of line between Exeter and Penzance also had to be altered from the old longitudinal timbers to the modern cross-sleepers (ties). It took an army of platelayers to do it, 4,200 of them (21).

#### Costs and Time

It is useful when planning the production and delivery of health services to consider the impact on the present value of costs of: (i) the starting date (postponement as already mentioned), (ii) the closing date (up to a possible date at which discounting renders the present value negligible), (iii) the rate of output (i.e., the output or outcomes expected per period), (iv) the volume of output (i.e., the cumulative output or outcomes between the starting and closing dates).

The faster the rate of an activity, in general the higher the present value of its cost for any planned program period. This is a standard expectation of most estimated cost functions, suggesting that marginal cost is positive. It may not hold universally, however, especially at relatively low rates of output when, say, newly created clinics are being rolled out and new patients recruited. Again, context matters and conventional empirical assumptions always need checking.

The greater the volume of activity the greater the present value of a program’s cost. The volume of a service is the cumulative output over a defined period, that is, a planned program. Plainly, given both the date at which a program begins and a constant rate of activity, enlarging its volume suggests a more distant endpoint. Alternatively, with a given (chosen) endpoint, volume may increase only if the rate of activity also rises.

The greater the volume of activity given a constant rate of output, the lower the marginal cost (i.e., the present value of cost rises but at a decreasing rate). The reason for this volume effect is usually that planning for larger volumes enables the use of more cost-effective methods.

For example, at low volumes patients may be referred elsewhere for some diagnostic tests, whereas a larger volume makes desirable the purchase and use of a scanner of one's own. The effect is often referred to as a scale economy. The larger the volume, the lower the present value of cost of each unit of output/outcome (average cost).

If both the rate and the volume of activity increase, the impact on marginal cost is ambiguous and is entirely context-dependent, that is, requiring specific evidence from trials, pilots, and the like.

### Social Cost, Externalities, and Publicness

An external cost is a cost imposed on those other than the decision maker. As with "internal" costs, one should distinguish between opportunity costs and harms. An external harm occurs when an infectious person communicates a disease to others having a direct negative impact on their welfare but not on their use as a resource. There may also be external opportunity costs, for example, in the form of defensive health care to mitigate the transmitted disease or reduction in productivity through absenteeism and presenteeism. External costs (of relevance in HTA) include the nonhealth sector consequences of hazards created by other people's behavior (e.g., in traffic) and pollution by industry and private households.

Social costs are the sum of internal (private) and external costs. It is unhelpful, as is sometimes found, to treat external costs as equivalent to social costs. External costs, when they exist, are an element of social cost along with internal costs. When they are absent, social cost is the same as internal cost.

External costs generally arise in the absence of appropriate property rights, specifically in the ability of people to be able to establish titles, for example, to a bug- and pollution-free environment, or its converse. Either way, with tradeable rights, either the imposer of the cost must compensate those on whom the costs are thrust or, in the converse case, the polluter must be compensated for reductions in pollution (22). Because many of these external effects have the character of publicness, in the technical sense that the effect, if it exists, impacts on many individuals who cannot be excluded from the effect. In such circumstances, there will often be a need for public regulation of tax/subsidy arrangements that, for example, subsidize or mandate vaccination programs (as when vaccination is a requirement before children may enter primary schools) or for markets in pollution rights.

The optimal level of pollution, like the optimal incidence or prevalence of a common disease, is unlikely to be zero. There are two main reasons for this. One is that the marginal

benefit falls as the external cost or harm is increasingly controlled, while the marginal cost of controlling it commonly rises. The optimal level will, therefore, typically be higher than zero because beyond the optimum the social value of further elimination is less than its social cost. In public health-care systems with a fixed budget, exceeding this optimal level suggests that more health is being sacrificed by spending more of the limited budget than further elimination will generate.

The other reason is that there are many other opportunities for public expenditure on health, and a policy of pollution elimination to benefit health is unlikely to commend itself if that expenditure might generate a greater health benefit by being applied to other health care activity. These are both examples where the opportunity cost of expenditure of healthcare resources on one program is the health lost from not spending it on another.

The economic "publicness" of the effects of these decisions has little to do with the ownership of resources, which may be either public or private. The critical characteristic of a public good or a public bad in economics is that the effect (cost or benefit) cannot be individualized. As immunization programs approach herd immunity, *everyone* benefits; as road traffic increases without regulation, *all* highway users are exposed to increased risk. The choice between the use of public or private agents will rest on other features: partly on their relative efficiency, partly on their context-specific characteristics like human resources practices, partly on the costs and enforceability of contracts and governance structures, and partly on general social conventions regarding the accepted roles of the private and public sectors.

In conclusion, it turns out that cost, or opportunity cost, is not quite the simple idea so often claimed for it. Cost certainly represents a forgone opportunity but what is forgone may be the monetary value of a physical entity or a nonmonetary measure of what is forgone. The latter becomes a cost when it is convenient to have a nonmonetary way of valuing forgone opportunities. For example, a new program in a healthcare system uses resources that have a monetary cost. But these resources have alternative uses within the healthcare system with health consequences that might be quantifiable. If quantified, the health forgone because of the new program is an alternative measure of its opportunity cost, which may be a more helpful way for decision makers to see it.

Opportunity cost is not necessarily captured in systems of financial accounting, even when expressed as a monetary value. Costs in accounts need always to be scrutinized when being used to proxy opportunity cost. Cost is sometimes perceived only subjectively in, for example, decisions about the allocation of resources within an organization. It is highly context-dependent. It is always the result of a decision but who the decision maker is, what the objective function is, and what discretion they are allowed are contextual. Likewise, what a decision commits the decision maker to may be highly

variable and is again context-dependent (the stability and comprehensiveness of a benefits package in a public healthcare system or national health service will usually have long-term commitments that have cost consequences). The decision's essential characteristics (dating, duration, rate, and volume) are all choice variables for the decision maker and those to whom the decision maker is accountable, all of whom define the scope of the decision and its other relevant characteristics. Few costs are technically fixed and whether a cost is treated as fixed for the purposes in hand is as matter of choice and context.

The major take-home lessons are that cost in decisions concerning efficient resource allocation ought always to be opportunity cost and that cost is always context-dependent or contingent. It is contingent upon the decision to be taken; the identity and objectives of the decision maker; the budgetary, technical, and political constraints imposed upon the decision maker; their permitted discretion within these constraints; their accountability to various stakeholders; and the judgments (value and scientific) of the decision maker as to the considerations deemed relevant in any particular context.

## CONFLICTS OF INTEREST

Dr. Culyer has nothing to disclose.

## REFERENCES

1. Alchian AA. *Economic forces at work*. Indianapolis: Liberty Press; 1977:273–299.
2. Ferraro PJ, Taylor LO. (2005) Do economists recognize an opportunity cost when they see one? A dismal performance from the dismal science. 2005. <http://jzuluaga.com/wp-content/uploads/CostoDeOportunidad.pdf>; doi:10.2202/1538-0645.1469 (accessed September 28, 2018).
3. Frank RH, Bernanke BS. *Principles of microeconomics*. New York: McGraw-Hill/Irwin; 2001.
4. Parkin M. Opportunity cost: A re-examination. *J Econ Educ*. 2016;47:12–22. doi:10.1080/00220485.2015.1106361.
5. von Wieser F. *Natural value*. London and New York: Macmillan; 1893.
6. Robbins LC. Remarks upon certain aspects of the theory of costs. *Econ J*. 1934;44:1–18.
7. Buchanan JM. *Cost and choice: An inquiry in economic theory*. Chicago: University of Chicago Press; 1969.
8. Wiseman J. *Cost, choice and political economy*. Cheltenham: Edward Elgar; 1989.
9. Robbins LC. *An essay on the nature and significance of economic science*. London and New York: Macmillan; 1932.
10. Campillo-Artero C, Puig-Junoy J, Culyer AJ. Does MCDA trump CEA? *Appl Health Econ Health Policy*. 2018. <https://doi.org/10.1007/s40258-018-0373-y> (accessed September 28, 2018).
11. Jevons WS. *The theory of political economy* (reprint of 1931 edition). Charlottesville: Ibis; 1871.
12. Marshall A. *Principles of economics*. London and New York: Macmillan; 1890.
13. Wicksteed PH. *The common sense of political economy (1910)*. London: Routledge and Kegan Paul; 1933.
14. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess*. 2015;19:1–503.
15. Culyer AJ. *Deliberative processes*. London: Office of Health Economics; 2009.
16. Wilkinson T, Claxton KP, Sculpher MJ, et al. The International Decision Support Initiative reference case for economic evaluation: An aid to thought. *Value Health*. 2016;19:921–928.
17. Morton A, Adler A, Bel D, et al. Unrelated future costs and unrelated future benefits: Reflections on NICE guide to the methods of technology appraisal. *Health Econ*. 2016;25:933–938.
18. HM Treasury. *The Green Book: Appraisal and evaluation in central government*. London: TSO; 2003 (updated 2018). [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/685903/The\\_Green\\_Book.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf) (accessed July 22, 2018).
19. National Institute for Health and Social Care. *Guide to the methods of technology appraisal*. London: NICE. <https://www.nice.org.uk/article/pmg9/chapter/foreword> (accessed July 22, 2018).
20. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: Combined and uneven development. *Med Decis Making*. 2007;28:21–32.
21. Smith LA. The broad gauge story. 1985. <http://lionels.orpheusweb.co.uk/RailSteam/GWRBroadG/BGHist.html> (accessed February 23, 2018).
22. Coase RH. The problem of social cost. *J Law Econ*. 1960;3:1–44.